2nd International Symposium on
Information Geometry and its Applications
December 12-16, 2005, Tokyo
Pages 000–000

INFORMATION GEOMETRY AND STATISTICAL INFERENCE

DÉNES PETZ[1]

Abstract. Variance and Fisher information are ingredients of the Cramér-Rao inequality. Fisher information is regarded as a Riemannian metric on a quantum statistical manifold and we choose monotonicity under coarse graining as the fundamental property. The quadratic cost functions are in a dual relation with the Fisher information quantities and they reduce to the variance in the commuting case. The scalar curvature in a certain geometry might be interpreted as an uncertainty on a statistical manifold. Information geometry has a surprising application to the theory of geometric mean of matrices.

1. The Cramér-Rao Inequality

The Cramér-Rao inequality belongs to the basics of estimation theory in mathematical statistics. Its quantum analog was discovered immediately after the foundation of mathematical quantum estimation theory in the 1960's, see the book [10] of Helstrom, or the book [11] of Holevo for a rigorous summary of the subject. Although both the classical Cramér-Rao inequality and its quantum analog are as trivial as the Schwarz inequality, the subject takes a lot of attention because it is located on the highly exciting boundary of statistics, information and quantum theory.

As a starting point we give a very general form of the quantum Cramér-Rao inequality in the simple setting of finite dimensional quantum mechanics. For $\theta \in (-\varepsilon, \varepsilon) \subset \mathbb{R}$ a statistical operator ρ_θ is given and the aim is to estimate the value of the parameter θ close to 0. Formally ρ_θ is an $n \times n$ positive semidefinite matrix of trace 1 which describes a mixed state of a quantum mechanical system and we assume that ρ_θ is smooth (in θ). Assume that an estimation is performed by the measurement of a self-adjoint matrix A playing the role of an observable. A is called locally unbiased estimator if

$$\frac{\partial}{\partial\theta} \mathrm{Tr}\, \rho_\theta A \Big|_{\theta=0} = 1\,. \tag{1}$$

This condition holds if A is an unbiased estimator for θ, that is

$$\mathrm{Tr}\, \rho_\theta A = \theta \qquad (\theta \in (-\varepsilon, \varepsilon)). \tag{2}$$

[1]Supported by the Hungarian grant OTKA T032662.

To require this equality for all values of the parameter is a serious restriction on the observable A and we prefer to use the weaker condition (1).

Example 1. *Let*

$$\rho_\theta := \rho + \theta B,$$

where ρ is a positive definite density and B is a self-adjoint traceless operator. A is locally unbiased when $Tr\, AB = 1$.

Example 2. *Let*

$$\rho_\theta := \frac{\exp(H + \theta B)}{Tr\, \exp(H + \theta B)}$$

with a density matrix e^H. Assume that $Tr e^H B = 0$. The Frechet derivative of e^H is $\int_0^1 Tr\, e^{tH} B e^{(1-t)H}\, dt$. Hence A is locally unbiased if

$$\int_0^1 Tr\, D^t B D^{1-t} A\, dt = 1 \qquad with \qquad D = e^H.$$

Let $\varphi_0[K, L] = \mathrm{Tr}\, \mathbb{J}_0(K)L$ be an inner product on the linear space of self-adjoint matrices. $\varphi_0[\,\cdot\,,\,\cdot\,]$ and the corresponding superoperator $\mathbb{J}_0$ depend on the density matrix ρ_0, the notation reflects this fact. When ρ_θ is smooth in θ, as already was assumed above, then

$$\frac{\partial}{\partial\theta} \mathrm{Tr}\, \rho_\theta B \Big|_{\theta=0} = \varphi_0[B, L] \tag{3}$$

with some $L = L^*$. From (1) and (3), we have $\varphi_0[A, L] = 1$ and the Schwarz inequality yields

$$\varphi_0[A, A] \geq \frac{1}{\varphi_0[L, L]}\,. \tag{4}$$

This is the celebrated inequality of Cramér-Rao type for the locally unbiased estimator. We want to interprete the left-hand-side as a sort of generalized variance of A. To do this it is useful to assume that

$$\varphi_0[B, B] = \mathrm{Tr}\, \rho_0 B^2 \quad \text{if} \quad B\rho_0 = \rho_0 B\,. \tag{5}$$

However, in the non-commutative situation the statistical interpretation seems to be rather problematic and we call this quantity quadratic cost functional.

The right-hand-side of (4) is independent of the estimator and provides a lower bound for the quadratic cost. The denominator $\varphi_0[L, L]$ appears to be in the role of Fisher information here. We call it quantum Fisher information with respect to the cost function $\varphi_0[\,\cdot\,,\,\cdot\,]$. This quantity depends on the tangent of the curve ρ_θ. If the densities ρ_θ and the estimator A commute, then

(6)

$$L = \rho_0^{-1} \frac{d\rho_\theta}{d\theta} \quad \text{and} \quad \varphi_0[L, L] = \mathrm{Tr}\, \rho_0^{-1} \left(\frac{d\rho_\theta}{d\theta}\right)^2 = \mathrm{Tr}\, \rho_0 \left(\rho_0^{-1} \frac{d\rho_\theta}{d\theta}\right)^2.$$

We want to conclude from the above argument that whatever Fisher information and generalized variance are in the quantum mechanical setting, they are very strongly related. In an earlier work [17, 18] we used a monotonicity condition to make a limitation on the class of Riemannian metrics on the state space of a quantum system. The monotone metrics are called Fisher information quantities in this paper.

Since the sufficient and necessary condition for the equality in the Schwarz inequality is well-known, we are able to analyze the case of equality in (4). The condition for equality is

$$A = \lambda L$$

for some constant $\lambda \in \mathbb{R}$. Therefore the necessary and sufficient condition for equality in (4) is

$$\dot{\rho}_0 := \frac{\partial}{\partial \theta}\rho_\theta \Big|_{\theta=0} = \lambda^{-1}\mathbb{J}_0(A)\,. \tag{7}$$

Therefore there exists a unique locally unbiased estimator $A = \lambda \mathbb{J}_0^{-1}(\dot{\rho}_0)$, where the number λ is chosen such a way that the condition (1) should be satisfied.

Example 3. *Continue the notation of Example 1. For the family* ρ_θ

$$A = \frac{B}{Tr\, B^2}$$

is a locally unbiased estimator and in the inequality (4) *the equality holds when* $\varphi_0[X, Y] = Tr\, XY$, *that is,* $\mathbb{J}_0$ *is the identity. If* $Tr\, \rho B = 0$ *holds in addition, then the estimator is unbiased.*

Example 4. *For the family* ρ_θ *in Example 2,*

$$A = \frac{B}{\int_0^1 Tr\, D^t B D^{1-t} B\, dt}$$

is a locally unbiased estimator and in the inequality (4) *the equality holds when* $\mathbb{J}_0(K) = \int_0^1 D^t K D^{1-t}\, dt$. *This estimator is typically not unbiased. When* $HB = BH$, *then we have a classical exponential family in the quantum formalism, and then* B *is unbiased.*

2. Coarse-Graining and Monotonicity

In the simple setting in which the state is described by a density matrix, a coarse-graining is an affine mapping sending density matrices into density matrices. Such a mapping extends to all matrices and provides a positivity and trace preserving linear transformation. A common example of coarse-graining sends the density matrix ρ_{12} of a composite system $1+2$ into the (reduced) density matrix ρ_1 of component 1. There are several reasons to assume completely positivity about a coarse graining and we do so.

Assume that ρ_θ is a smooth curve of density matrices with tangent $A := \dot\rho$ at ρ. The quantum Fisher information $F_\rho(A)$ is an information quantity associated with the pair (ρ, A), it appeared in the Cramér-Rao inequality above and the classical Fisher information gives a bound for the variance of a locally unbiased estimator. Let now α be a coarse-graining. Then $\alpha(\rho_\theta)$ is another curve in the state space. Due to the linearity of α, the tangent at $\alpha(\rho_0)$ is $\alpha(A)$. As it is usual in statistics, information cannot be gained by coarse graining, therefore we expect that the Fisher information at the density matrix ρ_0 in the direction A must be larger than the Fisher information at $\alpha(\rho_0)$ in the direction $\alpha(A)$. This is the monotonicity property of the Fisher information under coarse-graining:

$$F_\rho(A) \ge F_{\alpha(\rho)}(\alpha(A)) \tag{8}$$

Although we do not want to have a concrete formula for the quantum Fisher information, we require that this monotonicity condition must hold. Another requirement is that $F_\rho(A)$ should be quadratic in A, in other words there exists a nondegenerate real bilinear form $\gamma_\rho(A, B)$ on the self-adjoint matrices such that

$$F_\rho(A) = \gamma_\rho(A, A). \tag{9}$$

The requirements (8) and (9) are strong enough to obtain a reasonable but still wide class of possible quantum Fisher informations.

We may assume that

$$\gamma_\rho(A, B) = \mathrm{Tr}\, A \mathbb{J}_\rho^{-1}(B^*). \tag{10}$$

for an operator $\mathbb{J}_\rho$ acting on matrices. (This formula expresses the inner product γ_D by means of the Hilbert-Schmidt inner product and the positive linear operator $\mathbb{J}_\rho$.) In terms of the operator $\mathbb{J}_\rho$ the monotonicity condition reads as

$$\alpha^* \mathbb{J}_{\alpha(\rho)}^{-1} \alpha \le \mathbb{J}_\rho^{-1} \tag{11}$$

for every coarse graining α. (α^* stand for the adjoint of α with respect to the Hilbert-Schmidt product. Recall that α is completely positive and trace preserving if and only if α^* is completely positive and unital.) On the other hand the latter condition is equivalent to

$$\alpha \mathbb{J}_\rho \alpha^* \le \mathbb{J}_{\alpha(\rho)}\,. \tag{12}$$

We proved the following theorem in [17].

Theorem 2.1. *If for every density matrix ρ a positive definite bilinear form γ_ρ is given such that* (8) *holds for all completely positive coarse grainings α and $\gamma_\rho(A, A)$ is continuous in ρ for every fixed A, then there exists a unique operator monotone function $f : \mathbb{R}^+ \to \mathbb{R}$ such that $f(t) = tf(t^{-1})$ and $\gamma_\rho(A, A)$ is given by the following prescription.*

$$\gamma_\rho(A, A) = Tr\, A \mathbb{J}_\rho^{-1}(A) \qquad \textit{and} \qquad \mathbb{J}_\rho = \mathbb{R}_\rho^{1/2} f(\mathbb{L}_\rho \mathbb{R}_\rho^{-1}) \mathbb{R}_\rho^{1/2}\,,$$

where the linear transformations $\mathbb{L}_\rho$ and $\mathbb{R}_\rho$ acting on matrices are the left and right multiplications, that is

$$\mathbb{L}_\rho(X) = \rho X \qquad \textit{and} \qquad \mathbb{R}_\rho(X) = X\rho\,.$$

Instead of operator monotone functions, $\mathbb{J}_\rho$ can be described equivalently by means: $\mathbb{J}_\rho = \mathbb{L}_\rho m \mathbb{R}_\rho$, where m is a mean of positive operators.

Via the operator $\mathbb{J}_\rho$, each monotone Fisher information determines a quantity

$$\varphi_\rho[A, A] := \mathrm{Tr}\, A \mathbb{J}_\rho(A) \tag{13}$$

which could be called quadratic cost function. According to (12) this possesses the monotonicity property

$$\varphi_\rho[\alpha^*(A), \alpha^*(A)] \le \varphi_{\alpha(\rho)}[A, A]\,. \tag{14}$$

Since (11) and (12) are equivalent we observe a one-to-one correspondence between monotone Fisher informations and monotone quadratic cost functions. Any such cost function has the property $\varphi_\rho[A, A] = \mathrm{Tr}\,\rho A^2$ for commuting ρ and A. The examples below show that it is not so generally.

The analysis in [17] led to the fact that among all monotone quantum Fisher informations there is a smallest one which corresponds to the function $f_m(t) = (1 + t)/2$. In this case

$$F_\rho^{\min}(A) = \mathrm{Tr}\, AL = \mathrm{Tr}\,\rho L^2, \qquad \text{where} \qquad \rho L + L\rho = 2A. \tag{15}$$

For the purpose of a quantum Cramér-Rao inequality the minimal quantity seems to be the best, since the inverse gives the largest lower bound. In fact, the matrix L has been used for a long time under the name of symmetric logarithmic derivative, see [11] and [10]. In this example the quadratic cost function is

$$\varphi_\rho[A, B] = \tfrac{1}{2}\mathrm{Tr}\,\rho(AB + BA) \tag{16}$$

and we have

$$\mathbb{J}_\rho(A) = \tfrac{1}{2}(\rho A + A\rho) \qquad \text{and} \qquad \mathbb{J}_\rho^{-1}(A) = L = 2\int_0^\infty e^{-t\rho} A e^{-t\rho}\, dt \tag{17}$$

for the operator $\mathbb{J}$ of the previous section. (To see the second formula, set $A(t) := e^{-t\rho} A e^{-t\rho}$. Then

$$\frac{d}{dt} A(t) = -\rho A(t) - A(t)\rho$$

and

$$\rho \int_0^\infty A(t)\, dt - \int_0^\infty A(t)\, dt\, \rho = \big[- A(t) \big]_0^\infty = A.$$

This gives $\mathbb{J}_\rho^{-1}(A)$.)

Fisher information appears not only as a Riemannian metric but as an information matrix as well. Let $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$ be a smooth

m-dimensional manifold of invertible density matrices. The quantum score operators (or logarithmic derivatives) are defined as

$$L_i(\theta) := \mathbb{J}^{-1}_{\rho(\theta)}\big(\partial_{\theta_i}\rho(\theta)\big) \qquad (1 \le i \le m) \tag{18}$$

and

$$I^Q_{ij}(\theta) := \mathrm{Tr}\, L_i(\theta)\mathbb{J}_{\rho(\theta)}\big(L_j(\theta)\big) \qquad (1 \le i, j \le m) \tag{19}$$

is the quantum Fisher information matrix.

Theorem 2.2. [19] *Let α be a coarse-graining sending density matrices on the Hilbert space $\mathcal{H}_1$ into those acting on the Hilbert space $\mathcal{H}_2$ and let $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$ be a smooth m-dimensional manifold of invertible density matrices on $\mathcal{H}_1$. For the Fisher information matrix $I^{1Q}(\theta)$ of $\mathcal{M}$ and for Fisher information matrix $I^{2Q}(\theta)$ of $\alpha(\mathcal{M}) := \{\alpha(\rho(\theta)) : \theta \in G\}$ we have the monotonicity relation*

$$I^{2Q}(\theta) \le I^{1Q}(\theta). \tag{20}$$

The monotonicity of the Fisher information matrix in some particular cases appeared already in the literature: [16] treated the case of the Kubo-Mori inner product and [5] considered the symmetric logarithmic derivative and measurement in the role of coarse graining.

Assume that F_j are positive operators acting on a Hilbert space $\mathcal{H}_1$ on which the family $\mathcal{M} := \{\rho(\theta) : \theta \in G\}$ is given. When $\sum_{j=1}^n F_j = I$, these operators determine a measurement. For any $\rho(\theta)$ the formula

$$\alpha(\rho(\theta)) := \mathrm{Diag}\,(\mathrm{Tr}\,\rho(\theta)F_1, \dots, \mathrm{Tr}\,\rho(\theta)F_n)$$

gives a diagonal density matrix. Since this family is commutative, all quantum Fisher informations coincide with the classical (6) and the classical Fisher information stand on the left-hand-side of (20). The right-hand-side can be arbitrary quantum quantity but the most efficient is the symmetric logarithmic derivative.

3. The Cramér-Rao Inequalities Revisited

Let $\mathcal{M} := \{\rho_\theta : \theta \in G\}$ be a smooth m-dimensional manifold and assume that a collection $A = (A_1, \dots, A_m)$ of self-adjoint matrices is used to estimate the true value of θ.

Given an operator $\mathbb{J}$ we have the corresponding cost function φ_θ for every θ and the cost matrix of the estimator A is a positive definite matrix, defined by $\varphi_\theta[A]_{ij} = \varphi_\theta[A_i, A_j]$. The **bias** of the estimator is

$$\begin{aligned} b(\theta) &= \big(b_1(\theta), b_2(\theta), \dots, b_m(\theta)\big) \\ &:= \big(\mathrm{Tr}\,\rho_\theta(A_1 - \theta_1), \mathrm{Tr}\,\rho_\theta(A_2 - \theta_2), \dots, \mathrm{Tr}\,\rho_\theta(A_m - \theta_m)\big). \end{aligned}$$

For an unbiased estimator we have $b(\theta) = 0$. From the bias vector we form a bias matrix

$$B_{ij}(\theta) := \partial_{\theta_i} b_j(\theta).$$

For a locally unbised estimator at θ_0, we have $B(\theta_0) = 0$.

The relation

$$\partial_{\theta_i} \mathrm{Tr}\, \rho_\theta H = \varphi_\theta[L_i(\theta), H]$$

determines the logarithmic derivatives $L_i(\theta)$. The Fisher information matrix is

$$I_{ij}(\theta) := \varphi_\theta[L_i(\theta), L_j(\theta)].$$

Theorem 3.1. *Let $A = (A_1, \dots, A_m)$ be an estimator of θ. Then for the above defined quantities the inequality*

$$\varphi_\theta[A] \geq \big(I + B(\theta)\big) J(\theta)^{-1} \big(I + B(\theta)^*\big)$$

holds in the sense of the order on positive semidefinite matrices.

Concerning the proof we refer to [19].

4. Statistical Uncertainty

Assume that a manifold $\mathcal{M} := \{\rho_\theta : \theta \in G\}$ of density matrices is given together a statistically relevant Riemannian metric γ_d. We do not give a formal definition of such a metric. What we have in mind is the property that given two points on the manifold their geodesic distance is interpreted as the statistical distinguishability of the two density matrices in some statistical procedure.

Let $\rho_0 \in \mathcal{M}$ be a point on our statistical manifold. The geodesic ball

$$B_\varepsilon(\rho_0) := \{\rho \in \mathcal{M} : d(\rho_0, \rho) < \varepsilon\}$$

contains all density matrices which can be distinguished by an effort smaller than ε from the fixed density ρ_0. The size of the inference region $B_\varepsilon(\rho_0)$ measures the statistical uncertainty at the density ρ_0. Following Jeffrey's rule the size is the volume measure determined by the statistical (or information) metric. More precisely, it is better to consider the asymptotics of the volume of $B_\varepsilon(\rho_0)$ as $\varepsilon \to 0$. According to differential geometry

$$Vol\big(B_\varepsilon(\rho_0)\big) = C_n \varepsilon^n - \frac{C_n}{6(n+2)} \mathrm{Scal}\,(\rho_0)\varepsilon^{n+2} + o(\varepsilon^{n+2}), \tag{21}$$

where n is the dimension of our manifold, C_n is a constant (equals to the volume of the unit ball in the Euclidean n-space) and $Scal$ means the scalar curvature, see 3.98 Theorem in [8]. In this way, the scalar curvature of a statistically relevant Riemannian metric might be interpreted as the average statistical uncertainty of the density matrix (in the given statistical manifold). This interpretation becomes particularly interesting for the full state space endowed by the Kubo-Mori inner product as a statistically relevant Riemannian metric.

Let $\mathcal{M}$ be the manifold of all invertible $n \times n$ density matrices. The Kubo-Mori (or Bogoliubov) inner product is given by

$$\gamma_\rho(A, B) = \mathrm{Tr}\,(\partial_A \rho)(\partial_B \log \rho). \tag{22}$$

In particular, in the affine parametrization we have

$$\gamma_\rho(A,B)=\int_0^\infty \operatorname{Tr} A(\rho+t)^{-1}B(\rho+t)^{-1}, \tag{23}$$

see [16]. On the basis of numerical evidences it was conjectured in [16] that the scalar curvature which is a statistical uncertainity is monotone in the following sense. For any coarse graining α the scalar curvature at a density ρ is smaller than at $\alpha(\rho)$. The average statistical uncertainty is increasing under coarse graining. Up to now this conjecture has not been proven mathematically. (Some partial results were obtained in [2].) Another form of the conjecture is the statement that along a curve of Gibbs states

$$\frac{e^{-\beta H}}{\operatorname{Tr} e^{-\beta H}}$$

the scalar curvature changes monotonly with the inverse temperature $\beta \ge 0$, that is, the scalar curvature is monotone decreasing function of β.

5. Information Geometry on Positive Matrices

The positive definite matrices might be considered as the variance of multivariate normal distributions and the information geometry of Gaussians yields a natural Riemannian metric. The simplest way to construct an information geometry is to start with an information potential function and to introduce the Riemannian metric by the Hessian of the potential. We want a geometry on the family of non-degenerate multivariate Gaussian distributions with zero mean vector. Those distributions are given by a positive definite real matrix A in the form

$$f_A(x):=\frac{1}{\sqrt{(2\pi)^n \det A}}\exp\big(-\langle A^{-1}x,x\rangle/2\big) \qquad (x\in\mathbb{R}^n). \tag{24}$$

We identify the Gaussian (24) with the matrix A and have a simple and natural manifold structure. The tangent space at each foot point is the set of symmetric matrices.

There are many reasons (originated from statistical mechanics, information theory and mathematical statistics) that the Boltzmann entropy is a candidate for being an information potential. Up to some constants it is

$$S(f_A):=\log\det A=\operatorname{Tr}\log A\,. \tag{25}$$

By simple calculation we have

$$g_A(H_1,H_2):=-\frac{\partial^2}{\partial s\partial t}S(f_{A+tH_1+sH_2})\Big|_{t=s=0}=\operatorname{Tr} A^{-1}H_1A^{-1}H_2\,. \tag{26}$$

The corresponding information geometry of the Gaussians was discussed in [14] in details. We note here that this geometry has many symmetries, each similarity transformation of the matrices becomes a symmetry. In the statistical model of multivariate distributions (26) plays the role of the Fisher-Rao metric.

(26) determines a Riemannian metric on the set $\mathcal{P}$ of all positive definite complex matrices as well and below we prefer to consider the complex case. The geodesic connecting $A, B \in \mathcal{P}$ is

$$\gamma(t) = A^{1/2}(A^{-1/2}BA^{-1/2})^t A^{1/2} \qquad (0 \le t \le 1)$$

and we observe that the midpoint $\gamma(1/2)$ is just the geometric mean $A\#B$. The geodesic distance is

$$\delta(A, B) = \| \log(A^{-1/2}BA^{-1/2})\|_2 ,$$

where $\| \cdot \|_2$ stands for the Hilbert–Schmidt norm. (It was computed in [2] that the scalar curvature of the space $\mathcal{P}$ is constant.) These observations show that the information Riemannian geometry is adequate to treat the geometric mean of positive definite matrices [4, 13]. We arrived at a surprising application of the information geometry: The geometric mean of two (or more) positive matrices can be visualized in an information geometry.

References

[1] S. Amari *Differential-geometrical methods in statistics.* Lecture Notes in Stat. **28** Springer, Berlin, Heidelberg, New York, 1985.

[2] A. Andai Information geometry in quantum mechanics. Ph.D. dissertation (in Hungarian), BUTE, Budapest, 2004.

[3] O. E. Barndorff-Nielsen, R. D. Gill Fisher information in quantum statistics *J. Phys. A: Math. Gen.* **33**(2000), 4481–4490.

[4] R. Bhatia, J. Holbrook Geometry and means. preprint, 2005.

[5] S. L. Braunstein, C. M. Caves Statistical distance and the geometry of quantum states. *Phys. Rev. Lett.* **72**(1994), 3439–3443.

[6] N. N. Cencov Statistical decision rules and optimal inferences. *Translation of Math. Monog.* 53, Amer. Math. Society, Providence, 1982.

[7] A. Fujiwara, H. Nagaoka Quantum Fisher metric and estimation for pure state models. *Phys. Lett. A* **201**(1995), 119–124.

[8] S. Gallot, D. Hulin, J. Lafontaine *Riemannian geometry.* Springer, 1993.

[9] M. R. Grasselli, R. F. Streater On the uniqueness of the Chentsov metric in quantum information geometry. *Infinite Dimensional Anal. Quantum Prob.* **4**(2001), 173–182.

[10] C. W. Helstrom *Quantum detection and estimation theory.* Academic Press, New York, 1976.

[11] A. S. Holevo *Probabilistic and statistical aspects of quantum theory.* North-Holland, Amsterdam, 1982.

[12] A. Jenčova, D. Petz Sufficiency in quantum statistical inference. to be published in Commun. Math. Phys.

[13] M. Moakher A differential geometric approach to the geometric mean of symmetric positive definite matrices. *SIAM J. Matrix Anal. Appl.* **26**(2005), 735–747.

[14] A. Ohara, N. Suda and S. Amari Dualistic differential geometry of positive definite matrices and its applications to related problems. *Linear Algebra Appl.* **247**(1996), 31–53.

[15] M. Ohya, D. Petz *Quantum Entropy and Its Use*. Springer-Verlag, Heidelberg, 1993.

[16] D. Petz Geometry of Canonical Correlation on the State Space of a Quantum System. *J. Math. Phys.* **35**(1994), 780–795.

[17] D. Petz Monotone metrics on matrix spaces. *Linear Algebra Appl.* **244**(1996), 81–96.

[18] D. Petz, Cs. Sudár, Geometries of quantum states, J. Math. Phys. **37**(1996), 2662–2673.

[19] D. Petz Covariance and Fisher information in quantum mechanics. *J. Phys. A: Math. Gen.* **35**(2003), 79–91.

[20] E. P. Wigner, M. M. Yanase Information content of distributions. *Proc. Nat. Acad. Sci. USA* **49**(1963), 910–918.

Alfréd Rényi Institute of Matematics, Hungarian Academy of Sciences, POB 127, H–1364 Budapest, Hungary

E-mail address: petz@math.bme.hu

URL: http://www.math.bme.hu/~petz